DOCUMENT RESUME

ED 440 113 TM 030 716

AUTHOR Price, Cristofer; Karweit, Nancy
TITLE The "Prospects" Study of Educational Growth and Opportunity: Overview of Study Design, Analysis Methods and Results.
PUB DATE 1999-00-00
NOTE 30p.; Paper presented at the Biennial Meeting of the Society for Research in Child Development (Albuquerque, NM, April 15-18, 1999) and the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Compensatory Education; Disadvantaged Youth; Elementary Education; Longitudinal Studies; *Program Evaluation; *Research Design; Standardized Tests; Test Results
IDENTIFIERS *Elementary Secondary Education Act Title I; Impact Evaluation

ABSTRACT

In 1988, Congress mandated a national longitudinal evaluation of the effects of the Title I program on students' academic achievement and classroom behavior. This paper presents an overview of the study design, analysis methods, and impact results for student achievement. The analysis used four types of designs to assess the impact of Title I: (1) statistical controls to account for differences between participants and nonparticipants; (2) post-hoc creation of a control group; (3) adding of an initial test score to the model; and (4) repeated measurements of students who received services in some, but not all, years. Data sources included surveys of nearly 40,000 students, and surveys of parents, teachers, principals, and district-level staff. Student achievement outcomes were analyzed in three-level hierarchical linear models in which repeated observations were clustered within students, and students were clustered within schools. Results also indicate that students who participated in Title I had lower standardized test scores and were rated by their teachers as lower achieving and exhibiting poorer classroom behavior than others. Growth curves of Title I students were parallel to, but not below, those of their peers. Results indicate that where students start is highly predictive of their performance in later grades, and that Title I services did not bridge the performance gap between low-achieving students in high poverty schools and their more advantaged classmates. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 440 113

The *Prospects* Study of Educational Growth and Opportunity
Overview of Study Design, Analysis Methods and Results

Cristofer Price, Abt Associates Inc.
Nancy Karweit

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Cristofer Price

1

BEST COPY AVAILABLE

This paper was presented at the 1999 biennial meeting of the Society for Research in Child Development in Albuquerque, New Mexico

TM030716

The *Prospects* Study of Educational Growth and Opportunity
Overview of Study Design, Analysis Methods and Results

Cristofer Price, Abt Associates Inc.
Nancy Karweit

This paper was presented at the 1999 annual meeting of the American Education Research Association in Montreal, Canada.

Abstract

The Title 1 (formerly called Chapter 1) program, designed in the 1960's to improve academic opportunities in high-poverty schools, constitutes the largest outlay of federal funding for public elementary and secondary education. As such, in 1988, congress mandated a national, longitudinal evaluation, of the effects of the program on students' academic achievement and classroom behavior. The key issue addressed by the evaluation involves whether Title 1 results in improved academic performance and other outcomes for at-risk students. This paper presents an overview of the study design, analysis methods, and impact results for student achievement.

Practical limitations precluded the use of a study design involving random assignment to control and treatment groups. Four types of analysis designs are presented to assess the impact of Title 1 in the absence of random assignment. The first strategy used statistical controls to account for the differences between participants and non-participants. The second approach involved the post-hoc creation of a comparison group. The third approach built on the second by adding an initial test score to the model. Finally, the fourth approach focused on within-student Title 1 effects by examining the repeated measurements of students who received services in some but not all years.

Data sources included surveys of students, parents, teachers, principals and district-level staff. Additionally school records and standardized achievement tests were used. The sample included nearly 40,000 students from 365 schools. Data were collected from three student cohorts. Students in the 1st-grade cohort were administered four surveys/tests, starting in the fall of first grade, followed by measurements in three consecutive springs. Students in the 3rd-grade cohort were measured a total of four years, starting in the spring of 3rd grade and continuing each spring for three additional years. Data from the 7th-grade cohort included springtime measurements from 7th, 8th and 9th grades.

Student achievement outcomes were analyzed in three-level hierarchical linear models in which repeated observations were clustered within students, and students were clustered within schools. Separate growth curves were estimated and compared for groupings of students based on the number of years and pattern of receipt of Title 1 services. An iterative process of model building was used to identify salient predictors of student performance including student background and demographic characteristics, classroom context, and school level organization, size, and policy.

Results indicate that students who participated in Title 1 had lower standardized test scores, were rated by their teachers as lower achieving, and as exhibiting poorer classroom behavior than nonparticipating classmates. The growth curves of the Title 1 students were parallel, but below those of their peers. In addition, students who received more years of Title 1 services scored below those receiving less help. The observed lockstep pattern of similar growth over the years indicates that where students start out, relative to their peers, is highly predictive of their performance in later grades. And that Title 1 services were not sufficient to close the performance gap between low-achieving students in high poverty schools and their more advantaged classmates.

1. Introduction

The Title 1 program (formerly called Chapter 1[1]), designed in the 1960's to improve academic opportunities in high-poverty schools, constitutes the largest outlay of federal funding for public elementary and secondary education. As such, in 1988, congress mandated a national, longitudinal evaluation, of the effects of the program on students' academic achievement and classroom behavior. The mandate further stipulated that the study should be conducted throughout the country in urban, suburban, and rural areas and be of sufficient size and scope to assess and evaluate the effect of the program in all regions of the country. The key issue addressed by the evaluation involves whether Title 1 results in improved academic performance and other outcomes for at-risk students.

This paper presents an overview of the study design, analysis methods used for the *Prospects* study of Title 1 impacts on student achievement. In Section 2 the *Prospects* sample is described in terms of design, it's longitudinal structure, the grade spans and calender years represented, the data instruments, and sizes of analysis samples. Section 3 presents an overview of the analysis methods. This section includes a brief description of four strategies for assessing the impact of Title 1 on students, our approach to modeling the hierarchically structured data, choice of a functional form to represent growth, and how we defined variables to represent of Title 1 participation in the models. Section 4 presents an overview of results of basic growth patterns for students receiving various amounts of Title 1 services, and the results of four strategies for assessing the impact of Title 1 on student achievement.

2. Sample

2.1) Sample Design

In order to satisfy the congressional mandate, the sample was necessarily large (initially involving nearly 40,000 students in 365 schools).[2] It covered a broad grade span from 1st to 9th grade, and tracked the *same* students from 1991 to 1994. As shown in Exhibit 1, data were collected from three student grade cohorts:

- **1st-grade cohort**: students beginning 1st grade in the fall of 1991, who were tracked from entry into school through completion of the 3rd grade in the spring of 1994;
- **3rd-grade cohort**: students enrolled in the 3rd grade in the 1990-91 school year, who were tracked from the end of their 3rd grade year (spring 1991) through completion of the 6th grade in the spring of 1994; and

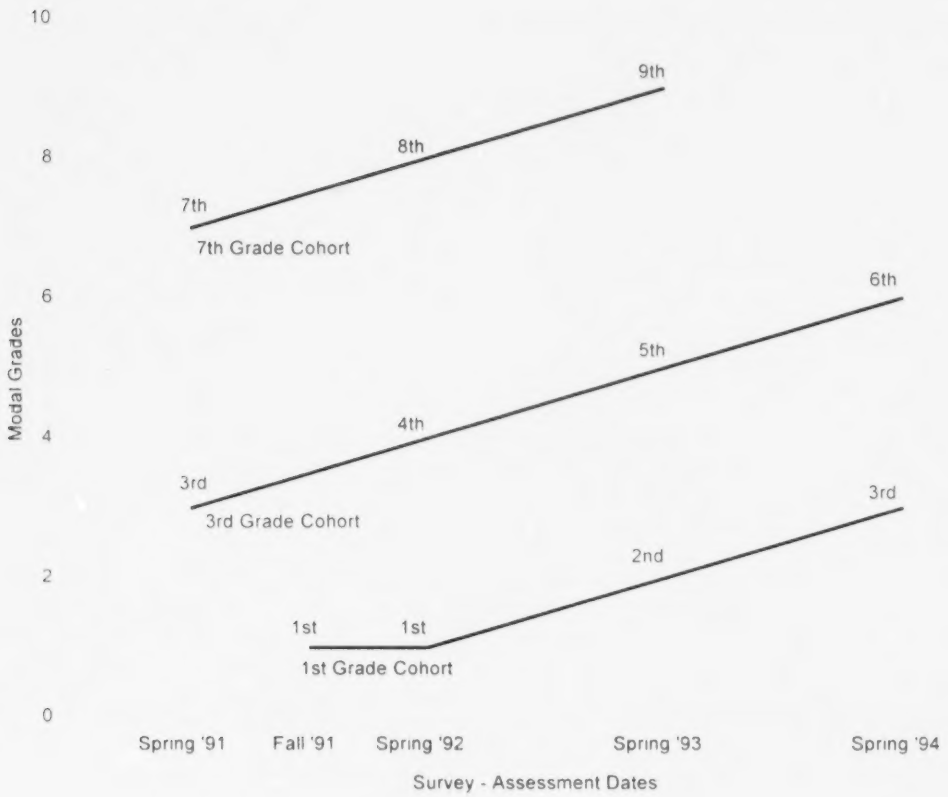
---

[1] The program was originally called Title 1. During the period covered by the *Prospects* study, 1990-1994, the program was called Chapter 1. The name subsequently changed back to Title 1. It will be referred to as Title 1 in this paper.

[2] Samples of students who moved from their initial school during this field period were followed and interviewed and tested in their new schools. By the end of the data collection period the total number of schools increased to 2,421.

- **7th-grade cohort**: students in the 7th grade during the 1990-91 school year, who were tracked from the end of 7th grade (spring 1991) through the completion of the 9th grade in the spring of 1993.

Exhibit 1: *Prospects* Data Collection Points by Study Cohort



Practical limitations precluded the use of a study design involving random assignment to control and treatment groups. Therefore, the main goal of the sample design was to ensure that the sample could support comparisons between:

- Title 1 participants,
- comparably disadvantaged children who did not have significant participation in the program, and
- their non-disadvantaged peers.

The sample design for *Prospects*[3] was implemented in three stages: 1) a sample of *school districts* was selected; 2) sample of *schools* within sampled districts was selected; and 3) where necessary[4], a sample of *students* in the designated grades within sampled schools was selected. Stratification was used to improve sample efficiency at each stage. Sample strata included four Census regions (northeast, south, midwest, west) and three urbanization categories (urban, suburban, rural).

2.2) Measures

The *Prospects* study collected a large amount of information from students (standardized tests and self-administered surveys), parents, teachers (both regular classroom and Title 1 instructors), principals, district-level staff, and students' school records. Data collected included the following:

- Student data were collected from several sources. The ***student abstract*** collected information from school records: absenteeism, tardiness, and grade retention; educational experiences such as preschool and Head Start attendance; identified disabilities; and participation in Title 1 or other remedial or special education programs. The ***student profile,*** completed by the teacher who was most knowledgeable about the student (usually the regular classroom teacher), included teacher ratings of the student on topics such as ability, motivation, attitudes, classroom socialization, language skills, and health status. Finally, a ***student questionnaire***, administered to students in the 3rd grade and above, collected data on current and past educational experiences, student and family demographics, educational aspirations, perceived academic strengths, grades, school attendance, participation in extracurricular activities within and outside of school, and parent participation in their schooling.

- The ***parent questionnaire*** provided information on parent and family characteristics, and additional information about parental attitudes and practices related to student learning. Topics included parental demographic information, household composition, income and employment, and home environment characteristics such as family rules, educational activities in the home, and parents' expectations for their child.

- ***Academic achievement*** was assessed using the Comprehensive Test of Basic Skills 4th edition, a vertically equated test series designed to measure

---

[3] The *Prospects* sample design is described in detail in a report produced for the *Prospects* Design Contract (Bryant, E.C., Chu, A. and Hansen, M. (1991) *Prospects: The Congressionally Mandated study of Educational Growth and Opportunity. Sample Design.* Produced under Department of Education Contract LC 89-027001.)

[4] Within most sampled schools, all students enrolled in all classrooms containing the target sample grades were included in the sample; only in schools with exceptionally large grade enrollments were students subsampled.

achievement in the basic skills taught in schools throughout the nation. The CTBS/4 was used to measure student achievement status and gains in reading/language arts and mathematics.

- The ***teacher instruments*** included information about teacher background, certification and experience, school climate (e.g., parent contacts, interactions with administration, and degree of influence over policies), and classroom instructional practices. Regular classroom teachers and Title 1 teachers were asked about the characteristics of their respective programs (teaching load, grade assignments, frequency of instruction, use of aides, instructional techniques), their responsibilities (subjects taught and extracurricular activities), their access to resources (including use of computer-assisted teaching programs), and their assessment of the performance level of their students. Teachers were also questioned about the coordination of the Title 1 program with regular teaching, and the selection process for admitting students to Title 1 instruction.

- The ***principal questionnaire*** and the ***characteristics of schools and programs*** *(CSP)* were both administered to the school principal or his/her designate. The principal questionnaire covered the following topics: the principal's background and experience, school policies, administrative techniques, and Title 1 program features (e.g., goals of the program, selection procedures, and parental involvement). The CSP questionnaire focused on the organization of the school (enrollment, grades taught, size of staff, class organization), demographics of the student body, and type and organization of Title 1 programs (and other special education programs) at each grade level.

- Finally, district information was obtained from the ***Title 1 district coordinator.*** Topics covered on this questionnaire included: district staff size, enrollment at each grade level, daily attendance, length of the school day, provision of in-service training for personnel, and demographics for the district (racial composition and poverty level). The district coordinator was asked about the characteristics of the Title 1 program, including the selection process for schools and students, instructional approaches in each subject area, staff size, method of evaluating the program, and activities for parental involvement. The district coordinator was also asked about characteristics of non-Title 1 compensatory education programs and services for language minority and Limited English Proficient (LEP) students.

2.1) Response Rate, Imputation, and Sample Sizes

Response rates for various survey instruments and achievements tests were typically above 80% for any particular wave of data collection (see *Prospects:* Final Report on Student Outcomes, Volume II Technical Report, Sept. 1997, for details). But, like any longitudinal

sample, proportions of students with full compliments of data across all years were considerably lower. Because there was such a great potential for analysis variables to be unavailable, and because data could not be assumed to be missing completely at random, an analysis of only those individuals with complete data is would have been inappropriate. Consequently, missing data imputation procedures were undertaken to obtain the best possible estimates of missing values, whenever possible. Details of the imputation process can be found in the Technical Report (Sept 1997). Below, we provide a brief summary of some of the more salient features of the imputation process.

- Imputations were done separately for four groups of variables: Outcomes, demographic/background variables, school variables, and teacher/classroom variables.
- Limits were set on the amount of missing data allowed before a variable could be included in the imputation procedure. Variables were included only if they had at least 50% of eligible cases. The actual incidence of missing data was generally much less than this cutoff level would suggest, however. Eligibility was defined as having had the opportunity to provide data.
- Title 1 participation variables were not imputed.

The analysis samples were created using those students who were contained in the original *Prospects* sample and who had the opportunity to be included in all waves of the survey for their respective cohort. There were some variations in the size of analysis samples dependent on the particular analysis. Exhibit 2 provides analysis sample sizes for each sample for the 1st and 3rd grade cohorts: N_1 represents the total number of observations included in the three level models, N_2 represents the total number of students in the longitudinal analyses, N_3 represents the number of schools that are included in the analyses, and N_4 represents the number of schools in the first year of the study.

The analyses described in "Longitudinal Patterns of Growth and Sources of Variation in Student Academic Achievement and Behavioral Outcomes", by Michael Vaden-Kiernan utilized the *student and school sample*. The analyses described in "Child, Family, Classroom, and School Predictors of Longitudinal Growth in Student Achievement" by Anne Ricciuti utilized the *student and school and classroom sample*. The *Title 1*, and *comparison groups* samples correspond to the Title 1 analyses described in the current paper. Exhibit 3 presents the various sample sizes of the subgroups used in the (subsequently described) within-subjects analyses.

Exhibit 2: Sample Information for *Prospects* Analysis Samples

Cohort	Sample	Subgroup	N_1	N_2	N_3	N_4
1	S1 Student and School	--	26,000	8,802	697	193
	S3, Student, School, Classroom	Math	24,350	8,585	677	192
		English	24,295	8,566	678	192
	S4: Title 1 Students	Math	22,671	7,631	530	190
		English	22,620	7,614	529	190
	S5: Comparison Group	Math	5,342	1,843	209	124
		English	6,203	2,144	226	128
3	S1 Student and School	--	35,768	9,218	1008	196
	S3, Student, School, Classroom	Math	33,267	8,939	956	195
		English	33,378	8,972	962	196
	S4: Title 1 Students	Math	29,818	7,619	803	191
		English	29,911	7,644	809	192

Exhibit 3: Sample Information for Within-Subject Analyses

Cohort	Sample	Subgroup	N_1	N_2	N_3
1	Math	1 Out of 3 Years	3,055	1,040	113
		2 Out of 3 Years	2,109	720	99
	Reading	1 Out of 3 Years	3,483	1,181	139
		2 Out of 3 Years	3,070	1,043	146
3	Math	1 Out of 4 Years	3,660	948	131
		2 Out of 4 Years	2,504	647	114
		3 Out of 4 Years	2,871	753	86
	Reading	1 Out of 4 Years	4,235	1,091	142
		2 Out of 4 Years	3,047	785	142
		3 Out of 4 Years	3,775	986	122

3) Overview of Analysis Methods

3.1) Finding A Comparison Group

The methodological question of paramount importance for this study is, "How do we know if Title 1 has had a positive impact on students?" The problem of attribution is at the heart of all research whether it be in physics, economics, or education policy. The problem of attribution in the *Prospects* study, was of student outcomes to special services rather than to other contributing factors (e.g., home environment). If governments, and program managers, are going to make difficult decisions about how much to invest in a particular policy initiative, or how to design and implement a service delivery program, they need to know with as much confidence as possible that their decisions are going to yield the desired results -- in this particular situation, improved school achievement for educationally disadvantaged children.

Over the last 30 years, social scientists have struggled with the challenge of how to determine whether a particular program contributed to a desired policy result, i.e., whether the outcomes observed for program participants (e.g., increased reading and math achievement) would have been achieved in the absence of the program. This is what is known as the problem of the counterfactual -- measuring what would have happened absent the program. In general, measuring changes for those who received some intervention is relatively easy, but measuring what would have occurred had they not received the assistance is not so simple.

The most widely accepted solution to this evaluation problem -- and the "gold standard" in both the physical and social sciences -- is the use of a randomized experiment. Under this research paradigm, eligible program participants are identified (either by their original eligibility

BEST COPY AVAILABLE

or through voluntary entry into the program) and are then randomly assigned to one of two groups. The first, or "treatment" group, is provided with the services or intervention that is being studied. In this case, this would mean that students would be identified as they are now to be in need of Title 1 assistance and they would receive such assistance as it is currently provided at their respective schools. The second, or "control" group, is identified as in need of services in exactly the same way (i.e., identification of the eligible pool must occur prior to the random assignment to the two study groups) but are ***not*** enrolled in the program. (These children may receive services at a later point, however.) By following the progress of both groups, and collecting the ***same*** information from both types of study members, we would be able to judge whether or not the program did, in fact, produce changes in the types of outcomes that are of interest, all other things equal. That is, a comparison of the average outcomes for the two groups would yield an unbiased and highly reliable, measure of the program's impact.

The use of random assignment evaluations has become nearly the expected standard social policy research in the U.S. and other industrialized nations, and in recent years has been increasingly evident in research studies in developing countries. Why is this the case? Simply because the gravity of the decisions that are being made -- billions of dollars in scarce resources expended and countless numbers of a society's most vulnerable people affected -- demands that we make decisions that are based upon information that is as accurate and reliable as possible.

Unfortunately, a randomized experiment was not implemented for this study, although it was one of the initially recommended options.[5] The design we have instead is a longitudinal study of selected children who do and do not receive Title 1 assistance, and who are in schools representing a complete spectrum of poverty concentration. As a consequence, absent the use of a randomized experiment, it is virtually impossible to establish a true natural comparison group, i.e., pupils who are exactly like those who receive Title 1 assistance but who are not served by the program. Districts select the most educationally disadvantaged children for services (using primarily the results of standardized achievement tests) in those schools that have the highest concentrations of poor children. Therefore, naturally occurring comparison groups will include either students in the same high-poverty schools who are measurably less educationally disadvantaged, or in a school that is not receiving Title 1 assistance (at least, in the same grade). On average then, a selection on non-participants will tend to produce a group of students who are marginally less disadvantaged.

To attempt to deal with this non-comparability, we have tried the four following statistical approaches (described in more detail later in this paper):

- **The first strategy used statistical controls to account for the differences between participants and non-participants.** (Title 1 participation variables were added to models with student, classroom, and school covariates).

- **The second approach involved the post-hoc creation of a comparison group** (a presumptively eligible group was defined as students in schools with 50% or more

---

[5] U.S. Department of Education(1989). *Planning Papers for the National Longitudinal Study of Title 1.* Washington DC.

eligible for free or reduced-price lunch, and who fell below the 25th percentile on the fall 1991 CTBS score).

- **The third approach built on the second by adding an initial test score to the model.**

- **Finally, the fourth approach focused on within-student Title 1 effects by examining the repeated measurements of students who received services in some but not all years.**

However, it is very difficult to attribute observed differences between participants and non-participants to the Title 1 program. That is, it is not clear whether observed differences are due to the effect of the program, or to underlying differences between the two populations. Statistical techniques are only as good as our ability to measure those characteristics that make the two groups of students different. And, even with perfect measurement tools, one is never certain that all the ways in which students can differ have been accounted for in the analytical model.

Fortunately, the "multiple" nature of the *Prospects* study design strengthens the use of statistical controls, i.e., data are available about *multiple* educational settings, for *multiple* cohorts of students, from *multiple* data sources, and over *multiple* years. However, the fact remains that conclusions about the effect of Title 1 that we draw from this study can only be considered suggestive -- conclusive attribution is simply not possible within the constraints of the study design.

3.2) Dealing With Hierarchically Structured Data

The approach taken in the analysis of the *Prospects* data attempted to capitalize on the hierarchical and longitudinal nature of the data to accurately model change in student outcomes over time. Multi-level analysis is appropriate when, as in the case of Title 1, outcomes are nested in a naturally occurring hierarchy of levels. In this situation, we focused on three levels of the hierarchy.[6] At the first level, we have repeated measures of student outcomes over time. These measures are nested within students, at the second level. At the second level we capture differences among individual students in their background and family characteristics, as well as in the types of instruction they receive. The third level consists of students nested within schools and captures variations across schools. With this type of analysis, effects can be modeled at each of these three levels. To give a specific example, we might consider a structure like that used by Bryk and Raudenbush (1992), considering, for exposition purposes, only between-student and

---

[6] It is conceptually possible to think of the *Prospects* data in terms of a 4-level model with individual observations at the first level, the repeated observations across students at the second level, students clustered within classrooms at the third level, and classrooms clustered within schools at the fourth level. We did, in fact, initially begin our analysis with this structure in mind. However, due to the fact that students changed classrooms every year, and the size and complexity of the data set, this type of 4-level model was not computationally feasible. Because there were relatively small numbers of student subjects clustered within any classroom, and there were large number of fixed effects describing classrooms, the effect, if any, of ignoring the classroom level of clustering on the standard error estimates may be expected to be small or negligible.

between-school level variation. At the within-student-level, outcomes are modeled as a student's individual growth trajectory; at the between-student-level, the student's "mean" response and growth parameters are modeled as a function of student background characteristics and the effect of the Title 1 program; at the between-school-level, mean school outcomes are modeled as a function of school policy, school atmosphere, and student body composition variables.

There are several advantages to such an approach. First, separate estimates of effects at each level of analysis avoids the problem of aggregation bias in the estimation of lower level covariate coefficients if the analysis is formulated at the higher level. This approach also yields efficient impact estimates because the non-independence of responses nested within groups is accounted for in the analytical procedure. In addition, multi-level modeling allows for the introduction of control variables at each level, giving the researcher greater ability to deal with potential biases. The main advantage of this approach, however, is that it provides the ability to model processes within students (over time), between students, and between schools simultaneously, with maximum statistical precision at each level. We therefore have a single analysis which models complex processes as they naturally occur.

The following discussion describes how we implemented this analytical approach. Throughout the discussion, the following general points should be kept in mind:

- Separate models were estimated for each of the **three study cohorts** -- 1st grade, 3rd grade, and 7th grade.
- Within grade cohorts, separate analyses were conducted for each of ten different student **outcome measures**: CTBS scale scores for Reading Vocabulary (SSRV); CTBS scale scores for Reading Comprehension (SSRC); CTBS scale score for Math Applications; CTBS scale score for Math Comprehension (SSMC); teacher rating of students' achievement in Reading (ACHVR); teacher rating of students' achievement in English (ACHVE); teacher rating of students' achievement in Math (ACHVM); teacher rating of students' classroom participation (PARTIC); teacher rating of students' cooperativeness (COOP); and teacher rating of students' attentiveness (ATTEN). The variables are defined as discussed earlier in this Title.
- Outcomes were modeled as a linear function of **grade**, **student** background covariates, **classroom** level covariates, **school** level covariates, and **Title 1 participation.**
- For the 1st grade cohort, the time points represented in the models are the spring time measurements from 1st, 2nd, and 3rd grades. The measures collected in the fall of first grade were used as a covariates in certain models.

3.3) Choosing an Appropriate Functional Form for "Growth"

The primary goals of our analysis were to: model student outcomes (i.e., test scores, behavior ratings, achievement ratings) as a function of increasing grade level; identify associations between covariates (e.g., student, school or classroom attributes) and the outcome measures and determine whether those associations vary with increasing grade level; and

identify associations between Title 1 participation and the selected student outcome measures. Consequently, one of the first issues that arose in developing the statistical models is the form of the underlying relationship between student outcomes and grade (i.e., time). Details of the exploratory process used for determining an appropriate functional form can be found in *Prospects*: Final Report on Student Outcomes, Volume II Technical Report (Sept. 1997). To summarize our decisions, we found that:

- For the first grade cohort, the rate of growth on CTBS/4 outcomes was much greater from first grade to second grade than from second to third grade. Modeling outcomes as a linear function of grade was therefore, only reasonable after finding appropriate transformations of the variables representing grade, and the outcome variables. For teacher ratings of behavior and achievement, no transformations were necessary.
- For the third and seventh grade cohorts, modeling outcomes as linear functions of grade appeared reasonable without the use of transformations.

3.4) Title 1 Participation Variables

The first approach to defining variables to represent Title 1 participation involved the specification of different patterns of participation, while a second approach (described in the following section) involved measures of the quantity or number of years the student received Title 1 assistance over the course of the study.

Title 1 Pattern Variables.

For the first grade cohort, eight Title 1 participation pattern variables were defined. The variables represented classification of students into one of the following eight categories according to their pattern of receipt of Title 1 services:

- students who never received Title 1 services (pattern000);
- students who received Title 1 services in 1992, but not 1993 or 1994 (pattern100);
- students who received Title 1 services in 1993, but not 1992 or 1994 (pattern010);
- students who received Title 1 services in 1994, but not 1992 or 1993 (pattern001);
- students who received Title 1 services in 1992 and 1993, but not 1994 (pattern110);
- students who received Title 1 services in 1992 and 1994, but not 1993 (pattern101);
- students who received Title 1 services in 1993 and 1994, but not 1992 (pattern011);
- students who received Title 1 services in 1992, 1993 and 1994 (pattern111).

Title 1 Amount Variables. An alternative classification scheme that was simpler than the pattern scheme was used to capture information on the quantity, or amount, of Title 1 assistance received by individual students. This step involved classifying students by the number of years of Title 1 services (either reading or math) they received over the observed period of time. Only the simpler "amount" scheme was used for students in the 3rd grade cohort because, for students who were followed all four years, the number of possible patterns of participation (16) was deemed to be unwieldy.

	Classification	Description of Classification
■	0 out of 4 years	student never received Title 1 services;
■	1 out of 4 years	pattern 1000 or pattern 0100 or pattern 0010 or pattern 0001;

- 2 out of 4 years — pattern 1100 or pattern 0110 or pattern 1010 or pattern 0101 or pattern 0011 or pattern 1001;
- 3 out of 4 years — pattern 1110 or pattern 1101 or pattern 1011 or pattern 0111;
- 4 out of 4 years — pattern 1111.

The "amount" classification scheme shown above corresponds to the third grade cohort. A similar scheme was also created for the first grade cohort.

4) Overview of Results

4.1) Growth Patterns and Title 1 Participation

We first overview a model of variation in student growth patterns by level of Title 1 participation. For the first grade cohort, the models were of the form:

$$Y_{ijk} = \beta_0 + \beta_1(\text{GRADE}_i) + \Sigma\, \beta_{m\cdot 1}(\text{Pattern}_j) + \Sigma\, \beta_{m\cdot p\cdot 1}(\text{GRADE}_i)(\text{Pattern}_j) + v_{0k} + v_{1k}(\text{GRADE}_i) + u_{oj} + u_{1j}(\text{GRADE}_i) + \varepsilon_{ijk},$$

where, Y_{ijk} is the outcome measure (transformed in models of 1st grade cohort CTBS scores) for student j in school k, the β's are fixed effects coefficients, GRADE (transformed in models of 1st grade cohort CTBS scores) represents student's grade in school. Pattern 000 was taken as the baseline and compared to the seven remaining patterns. Σ's represent summation over $m = 1...7$, the variable PATTERN is a (0,1) indicator variable indicating whether or not the j^{th} student is a member of the group classified by the m^{th} pattern, the v's represent random effects coefficients for the third level (schools), the u's represent random coefficients for the second level (students), and ε_{ijk} represents the random error at the first student level.

Exhibit 4 shows the predicted growth curves of Reading Vocabulary achievement for the first grade cohort. The patterns vary somewhat for other outcome measures, and for the other cohorts, but there are three striking observations that emerge from essentially all of the models. First, it is quite apparent that Title 1 is targeting children who are most in need of assistance. Across all of the different measures, Title 1 participants consistently score (or are rated) well below non-participating students. These differences are quite large. For example, the difference between the 1st grade Reading Vocabulary scores of non-participants and children who receive reading assistance in all three years is 66 scale score points (a difference of more than 1.5 standard deviations). These differences do not, however, necessarily indicate that Title 1 is not helping students (though, in part, they may reflect that result). What these findings certainly do show is that, on average, students who are most in need of help are the ones who are getting Title 1 services from year to year. Second, those students who receive more help (i.e., more years of Title 1 participation) score (or are rated) lower than those students who receive less help. Again, this is not a direct indicator of program impact, but rather it demonstrates the differences in performance between students who do and do not receive various levels of Title 1 assistance.

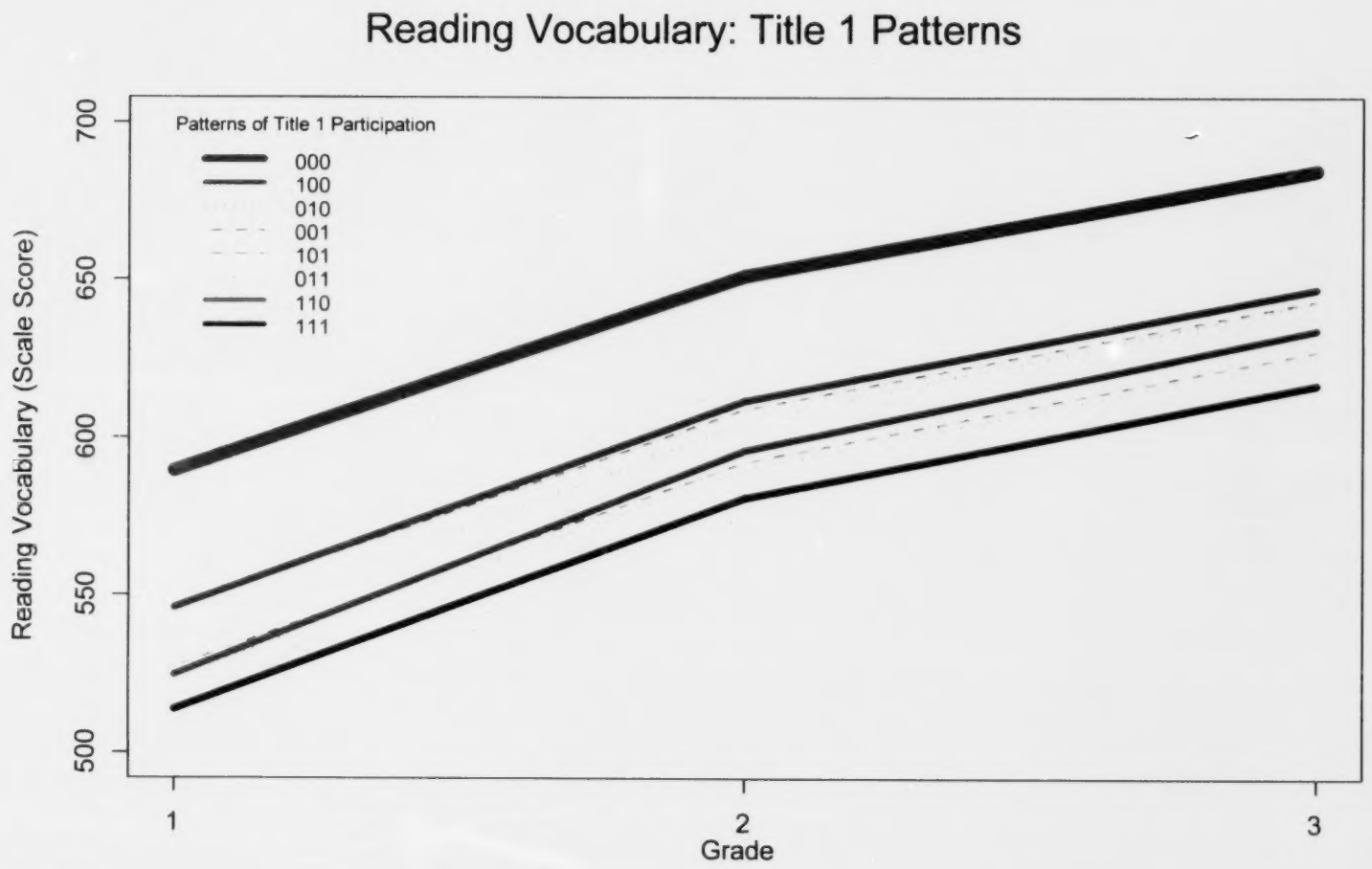
It is also worth noting that the position of the different groups of students remains the same over levels, particularly for the gap between the non-participants (the "000" group) and

those students who receive Title 1. Apparently, Title 1 services are not able to close the gap between participants and non-participants.

The observed stability of the performance gap does not necessarily mean that Title 1 is ineffective. Because of the design of the *Prospects* study we do not know how participants would have done in the absence of the program. It may be the case that without program services, the targeted students would have fallen further behind their grade-level peers (i.e., the gap may have widened). In fact, our expectation is that these students would fall farther behind (i.e., the gap would increase over time). But all we can say at this point is that where children enter school--as measured by standardized tests or teacher ratings--largely determines their relative position as they progress through school. If there is a failure in the program, it is its inability to "compensate" for differences in students' ability to meet expected levels of school performance.

This finding is not unexpected. Program participants were disadvantaged, relative to their peers, in a variety of ways that can affect the skills and abilities they bring to school. Furthermore, the intervention that most participants receive (which is related to the average student outcome) is relatively weak and unlikely to compensate for the observed achievement differences at the point of school entry.

Exhibit 4:


Reading Vocabulary: Title 1 Patterns
Patterns of Title 1 Participation
000
100
010
001
101
011
110
111
Reading Vocabulary (Scale Score)
500
550
600
650
700
1
2
3
Grade

4.2) Models using statistical controls to account for the differences between participants and non-participants.

The first statistical approach involved adding variables to control for student, family, school, and classroom characteristics to the models with measures of different dosages of Title 1 program assistance discussed in the previous section. If we were able to completely control for the differences between participants and non-participants using the covariates in our model, then the influence of the dosage variable would show the true impact of Title 1 services. In this scenario, we would hypothesize that students who received more help should do better than those who receive less. That is, if all other things are held equal by the model, getting more help should produce greater gains in the selected outcome measures than receiving less help.

Exhibit 5 provides the model results of the "amount plus covariates" models by grade cohort and outcome measure[7]. For each grade cohort, results are first presented for the model ***without*** controls for student, family, school, and classroom characteristics (labeled the "base model"--this model was discussed in Section 4.1) and then for the "fully specified" model which includes all the factors discussed in the paper by Riccuiti. For each model, the parameter estimates are shown for both the level (i.e., differences in average outcomes) and slope (i.e., differences in rates of change over time) effects. The cell entries represent the difference between the particular subgroup of students (i.e., 1st grade cohort with one year of Title 1 assistance) and students who did ***not*** receive Title 1 assistance in any of the observed years. For example, the first entry for the 1st grade cohort of "-34" means that students who received Title 1 assistance for one year had spring of 1st grade test scores that were, on average, 34 scale-score points in Math Applications ***below*** the average of those students who never received this type of Title 1 assistance between the 1st and 3rd grades.[8] Also provided in Exhibit 5 for each cohort is the percent reduction in this "gap" that resulted from adding all the covariates to the model (i.e., the percentage difference between the base and full models).

Starting with the base model, Exhibit 5 shows that across all outcome measures students who receive ***any*** Title 1 assistance have lower achievement scores, and lower teacher ratings, than students who do not receive this type of help in school. Moreover, those students who received more assistance consistently score below those students who received less (i.e., fewer years of) assistance. This does not necessarily indicate a negative impact of the program. Rather,

---

[7] The results shown correspond to the models with Title 1 participation parameterized as "amount" variables. The models with "pattern" variables produced qualitatively similar results.

[8] The standard deviations for the CTBS tests range from 40-50 scale-score points.

Exhibit 5: Summary of Statistical Model Results With Title 1 Amount Variables by Grade Cohort: Base Model vs. Fully-Specified Model

Outcome and Years of Title 1 Participation	1st Grade Cohort					3rd Grade Cohort				
	Base Model		Full Model			Base Model		Full Model		
	Level	Slope	Level	Slope	Percent Change	Level	Slope	Level	Slope	Percent Change
Math Applications Scale Score										
1	-34	N.S.	-29	N.S.	-15%	-27	N.S.	-22	N.S.	-19%
2	-35	N.S.	-28	N.S.	-20%	-45	2.3	-37	N.S.	-18%
3	-36	N.S.	-27	N.S.	-25%	-47	N.S.	-39	N.S.	-17%
4	--	--	--	--	--	-58	5.0	-46	4.8	-21%
Reading Vocabulary Scale Score										
1	-41	N.S.	-28	N.S.	-7%	-20	N.S.	-17	N.S.	-15%
2	-57	N.S.	-51	N.S.	-11%	-31	-3.0	-24	-4.6	-23%
3	-66	N.S.	-62	N.S.	-6%	-39	-2.7	-30	-3.4	-23%
4	--	--	--	--	--	-47	-2.4	-34	N.S.	-28%
Teacher's Math Achievement Rating										
1	-0.24	-0.10	-0.19	-0.08	-21%	-0.34	-0.05	-0.27	N.S.	-21%
2	-0.39	-0.12	-0.32	-0.10	-18%	-0.61	N.S.	-0.50	N.S.	-18%
3	-0.42	N.S.	-0.35	N.S.	-17%	-0.63	N.S.	-0.56	N.S.	-11%
4	--	--	--	--	--	-0.65	N.S.	-0.56	N.S.	-14%
Teacher's Reading Achievement Rating										
1	-0.39	-0.08	-0.32	N.S.	-18%	-0.44	N.S.	-0.36	N.S.	-18%
2	-0.60	-0.11	-0.50	-0.06	-17%	-0.71	N.S.	-0.54	N.S.	-24%
3	-0.64	-0.13	-0.57	-0.09	-11%	-0.81	N.S	-0.61	N.S.	-25%
4	--	--	--	--	--	-0.88	N.S.	-0.66	N.S.	-43%

BEST COPY AVAILABLE

Outcome and Years of Title 1 Participation	1st Grade Cohort					3rd Grade Cohort				
	Base Model		Full Model			Base Model		Full Model		
	Level	Slope	Level	Slope	Percent Change	Level	Slope	Level	Slope	Percent Change
Reading Teacher's Attention Rating*										
1	-0.22	-0.04	-0.17	N.S.	-23%	-0.22	N.S.	-0.18	-0.007	-18%
2	-0.33	N.S.	--0.30	N.S.	-9%	-0.38	0.04	-0.29	0.03	-24%
3	-0.30	N.S.	-0.34	N.S.	-13%	-0.36	N.S.	-0.30	N.S.	-17%
4	--	--	--	--	--	-0.26	N.S.	-0.29	N.S.	+11%
Math Teacher's Attention Rating **										
1	-0.24	-0.03	-0.12	-0.03	-50%	-0.22	N.S.	-0.18	N.S.	-18%
2	-0.40	N.S.	-0.19	N.S.	-53%	-0.36	0.03	-0.32	0.009	-11%
3	-0.41	N.S.	-0.14	N.S.	-66%	-0.38	N.S.	-0.33	N.S.	-13%
4	--	--	--	--	--	-0.37	N.S.	-0.22	N.S.	-41%

* Title 1 Participation in Reading.
** Title 1 Participation in Math
N.S. Not statistically significant at the 0.001 level.

BEST COPY AVAILABLE

it is probably just an indication that the right students are being targeted for help.

It is also the case that with few exceptions there are no significant slope effects, and where they are found, the slope effects are often negative.[9] This means that either there are no differences in growth (i.e., relative change over time), or where there are differences they tend to increase the observed gaps in average performance. What we would have expected to see is a ***positive*** effect on students' rate of growth (i.e., the slope effect) indicating that Title 1 assistance is helping close the gaps in student achievement. Again, because of the program's inherent selection bias we cannot conclude that the program is having no effect -- we simply do not know what would have happened to the participants in the absence of the program, since they are so different from participants and none of those differences have been removed. What we can say, however, is that the help they receive is generally not allowing them to move toward parity with their more advantaged peers (i.e., we found a non-significant slope effect).

What happens when statistical controls are added to the model? As shown in Exhibit 5, the magnitude of the level differences between participants and non-participants is reduced with the addition of the control variables but does not completely disappear. Participants still consistently score below non-participants. Furthermore, the differences between less and more assistance, while generally reduced, still persist in their stratified pattern, i.e., students with the most years of help score at the lowest level. In addition, the previously observed slope effects either become statistically insignificant or are reduced in size with the addition of the statistical controls, but do not become positive. That is, the negative gap between participants and non-participants continues over time.

The models we have estimated control for a large number of variables and do reduce the gap between the different groups of students. But, because there is essentially no effect on the rates of growth when control variables are added, the trajectories remain parallel. In other words, students who enter either the 1st or 3rd grade and are identified as in need of Title 1 assistance score, or are rated, below those that are not targeted for assistance. And, those that are targeted for help in multiple years score below those selected for less assistance. All students, whether they do or do not receive assistance, exhibit change over time. But, the rates of growth of Title 1 and non-Title 1 students are similar, indicating that Title 1 assistance does not "compensate" for the initial deficiencies of the disadvantaged students.

4.3) Title 1 Comparison Group Models.

In the models described above, outcome measures for students who received no Title 1 services were compared to those of students who received various amounts or patterns of Title 1 services, controlling for a wide range of student, family, school, and classroom covariates. In addition to this approach, we also tried to create a quasi-experimental, post-hoc comparison group to compare the average growth trajectories of 1st grade cohort students who received Title 1 services, to the trajectories for students in the same grade who would be eligible targets for Title 1 assistance but who, in fact, did not receive services.

For these analyses we used two criteria to determine "presumptive eligibility" for Title 1 services: the student must 1)have been enrolled in a school with greater than 50 percent of the students eligible for

---

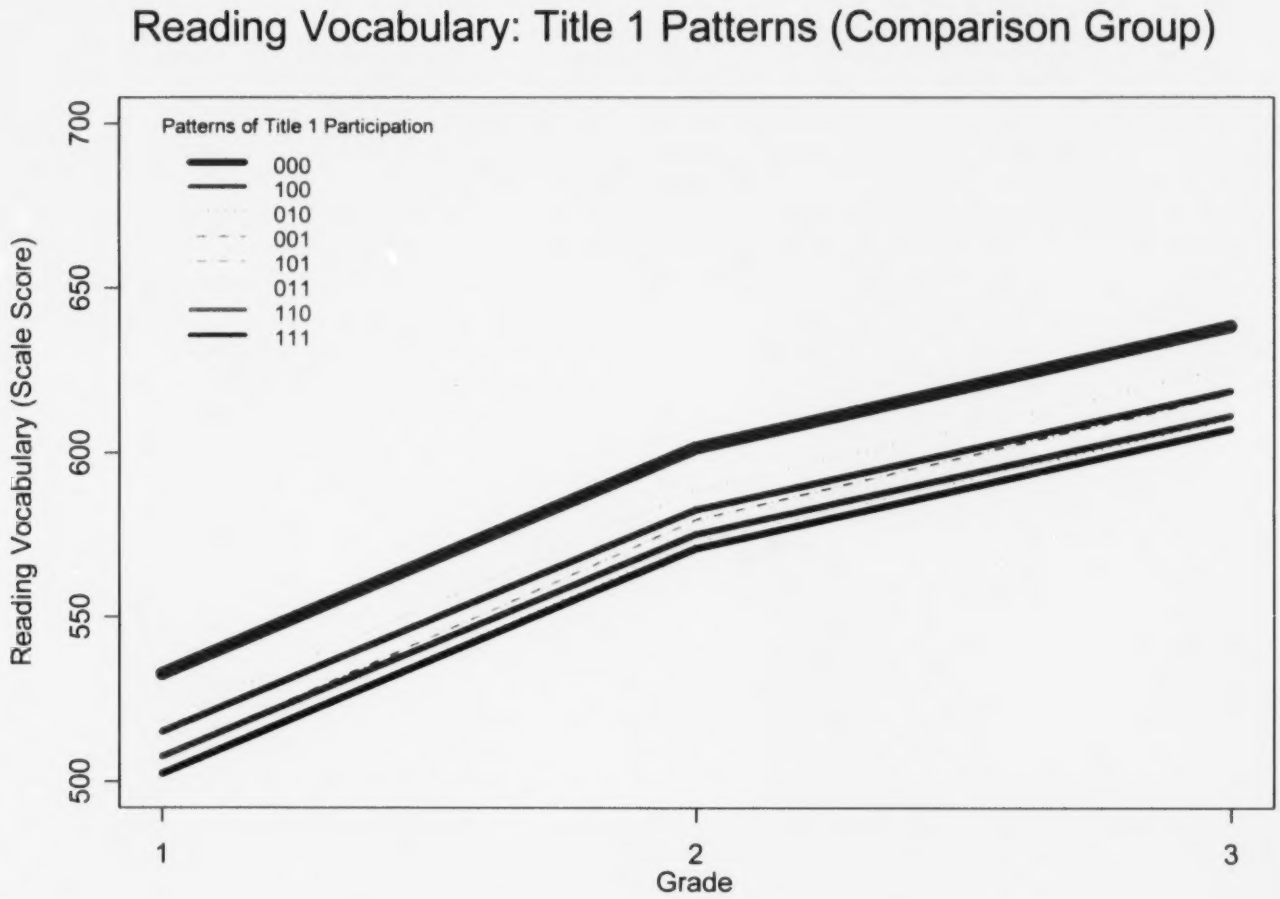
[9] Where numbers are shown in Exhibit 5 the differences were statistically significant at the 0.0001 level. We have used this more stringent criterion due to the large number of statistical tests that are being run.

free and/or reduced-price school meals, and 2) must have had a Fall 1991 test score lower than the 25th percentile based on national norms. Thus, for these analyses, to be classified as presumptively eligible for Title 1 reading services a student must have had a Fall 1992 CTBS Reading Vocabulary (SSRV) score below 452; presumptive eligibility for math services corresponded to fall 1992 CTBS score on Math Applications (SSMA) below 434. Data for students not meeting these eligibility criteria were excluded from this stage of the analysis.

For this sample of students, we re-estimated the previously described models for both Title 1 pattern and amount measures. A model for Reading Vocabulary was estimated using the sample of 1st grade cohort students who were eligible for reading services according to the criteria specified above; the same set of models was fit for Math Applications using the same rationale. The models included the same set of student, school, and classroom variables that came out of the previous analytical steps.

Exhibit 6 shows the predicted curves resulting from the estimated models for Reading Vocabulary. Because of our focus on a selected group of low-achieving and economically disadvantaged students, these curves are, on average, lower than those presented earlier. However, the results are essentially unchanged. That is, all students show growth over time, but students who receive Title 1 assistance score significantly below those who do not, and, those who receive more help score lower than those who receive less. Moreover, because the slope effects are again not significant, the different groups of students have the same growth rate, i.e., there is no "compensatory" effect over time associated with program participation.

Exhibit 6.


Reading Vocabulary: Title 1 Patterns (Comparison Group)
Patterns of Title 1 Participation
000
100
010
001
101
011
110
111
Reading Vocabulary (Scale Score)
500
550
600
650
700
1
2
3
Grade

4.4) Title 1 Comparison Group Models with Fall 1992 Predictor.

The analysis described above for the creation of a "presumptively eligible" group of 1st grade cohort students was replicated with the inclusion of a covariate for each student's initial Fall 1992 test score. This was an attempt to obtain further similarity between the two groups of students. All the other modeling specifications were as described above.

Adding the student's initial test score to the three-level model somewhat improved the comparability of the groups of participants and non-participants. But, despite this improvement the results are essentially the same as those discussed above. The only difference is for Math Applications where we find a small, but statistically significant, interaction between grade (i.e., time) and initial test score at the entry to 1st grade. In other words, students who scored lower at entry to school, grow at a slightly slower rate than other students.

The basic conclusion here is that the effect of Title 1 participation is independent of students' growth rates, i.e., there are still no positive slope differences between participants and non-participants.

4.5) Within-Subjects Analysis of Title 1 Participation.

A final attempt to examine the effect of Title 1 participation on student outcomes involved an examination of those students who received assistance in some but not all years. The question of interest here was, "For those students who received intermittent services, were there any systematic associations between test scores in years when services were received, versus years when no Title 1 services were received?"

To answer this question a series of models was estimated in which a time-varying Title 1 (0,1) indicator variable was included which took the value "1" if a student received services in the current year, and "0" otherwise. These models were estimated for the following subgroups of students:

- 1st grade cohort,
  - -- those who received assistance in one out of three years and,
  - -- those who received assistance in two out of three years.

- 3rd grade cohort,
  - -- those who received assistance in one out of four years;
  - -- those who received assistance in two out of four years; and,
  - -- those who received assistance in three out of four years.

A positive estimate for the Title 1 indicator variable in this context would indicate that, on average, students scored higher (relative to their own growth curves) during the years they received Title 1 services than the years they did not receive services.

The models were of the following form:

$$Y_{ijk} = \beta_0 + \beta_1(\text{GRADE}_i) + \beta_2(\text{Title1}_i) + v_{0jk} + v_{1k}(\text{GRADE}_i) + u_{oj} + u_{1j}(\text{GRADE}_i) + \varepsilon_{ijk},$$

where, Y_{ijk} is the outcome measure (or the transformed 1st grade cohort CTBS scores) at grade 1 for student j in school k, the β's are fixed effects coefficients, GRADE is coded as above, (Title1_i) is a (0,1) variable indicating whether a student received Title 1 services at GRADE_i, the v's represent random effects coefficients for the third level (schools), the u's represent random coefficients for the second level (students), and ε_{ijk} represents the random error at the first student level.

The models were estimated for Reading Vocabulary and Math Applications under two different scenarios: 1) students were allowed to have their own levels and growth curves (i.e., slopes); and 2) students were fixed at average slopes but allowed to have varying individual levels.

For both modeling approaches, the results are essentially the same, i.e., we did ***not*** detect any statistically significant differences from the student-specific expected yearly growth in test scores between the in which students did and did not receive assistance. These results, then, support what was found in earlier analyses to date, there is no indication in the *Prospects* data that Title 1 participation improves students' educational outcomes. Later research, if able to overcome the inherent selection bias, may yield such evidence.

5. Summary

The initial descriptive analysis of the patterns of student achievement for those children who did and did not receive Title 1 assistance indicated a lack of a "compensatory effect." That is, disadvantaged children exhibited growth over time, but they were seemingly unable to close the gap in achievement between themselves and their nondisadvantaged peers.

The use of statistical techniques to extend this analysis by controlling for measurable differences between participants and nonparticipants provided a more valid basis from which to judge program effects. The key finding from these analyses was that controlling for a variety of student, family, school, and classroom variables did not change the basic descriptive conclusions:

- Students selected to receive Title 1 assistance had test scores, or were rated by their teachers, below their nonparticipating classmates;
- Those students who received more Title 1 assistance (i.e., more years of service) had test scores, or were rated, below those who received less help; and,
- Title 1 students did not, over time, close the performance gap between themselves and their more advantaged classmates.

The first implication to be drawn from these results is that Title 1 was, on average, serving those students who were clearly most in need of supplementary assistance. That is, students who received assistance scored below those who did not, and repeated selection for compensatory education was associated with greater educational deficits.

Second, Title 1 assistance was, on average, insufficient to close the gap, or compensate for initial achievement differences between advantaged and disadvantaged students. The observed "lockstep pattern" of student growth clearly demonstrated that where students started out, relative to their classmates, is where they ended up in later grades.

Finally, our inability to discern a compensatory effect of Title 1 is not necessarily an indication of program failure. As noted above, the design of the *Prospects* study does not allow us to observe directly whether Title 1 students would have been worse off (i.e., whether the gap would have widened over time) in the absence of the services they received. In fact, because we might expect the gaps to grow over time, without a special intervention, it may be that Title 1 is helping participating students but is too weak an intervention to bring them up to par with their classmates.

Final Reports

Puma, M.J., Karweit, N., Price, C. Ricciuti, A., Thompson, W., Vaden-Kiernan, M., (1997). "Prospects: Final Report on Student Outcomes." Prepared for Planning and Evaluation Service, U.S. Department of Education.

Puma, M.J., Karweit, N., Price, C., Ricciuti, A., Vaden-Kiernan, M., (1997). "Prospects: Final Report on Student Outcomes - Volume II: Technical Report." Prepared for Planning and Evaluation Services, U.S. Department of Education.